



Colocation and Scientific Collaboration: Evidence from a Field Experiment

Citation

Boudreau, Kevin, Ina Ganguli, Patrick Gaule, Eva Guinan, and Karim Lakhani. "Colocation and Scientific Collaboration: Evidence from a Field Experiment." Harvard Business School Working Paper, No. 13-023, August 2012

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:9502859>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)



Colocation and Scientific Collaboration: Evidence from a Field Experiment

**Kevin Boudreau
Ina Ganguli
Patrick Gaule
Eva Guinan
Karim Lakhani**

Working Paper

13-023

August 22, 2012

Copyright © 2012 by Kevin Boudreau, Ina Ganguli, Patrick Gaule, Eva Guinan, and Karim Lakhani

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

Colocation and Scientific Collaboration: Evidence from a Field Experiment

By Kevin Boudreau, Ina Ganguli, Patrick Gaule,
Eva Guinan and Karim Lakhani*

August 22, 2012

We present the results of a field experiment conducted within the Harvard Medical School system of hospitals and research centers to understand how colocation impacts the likelihood of scientific collaboration. We introduce exogenous colocation and face-to-face interactions for a random subset of biomedical researchers responding to an opportunity to apply for a research grant. While the overall baseline likelihood of any two researchers collaborating is small, we find that random colocation significantly increases the likelihood of pair-level co-application by almost 70%. The effect of exogenous colocation on subsequent collaboration was greater for previous coauthors, pairs including a woman, and pairs researching similar clinical areas. Our results suggest that matching between scientists may be subject to considerable frictions—even among those in relatively close geographic proximity and in the same organizational system. At the same time, even a brief and focused intervention facilitating face-to-face interactions can provide information that impacts the formation of scientific collaborations.

* Kevin J. Boudreau, London Business School & Harvard Institute for Quantitative Social Science (kboudreau@london.edu); Ina Ganguli, SITE, Stockholm School of Economics (ina_ganguli@hksphd.harvard.edu); Patrick Gaule, CERGE-EI (pgaule@cerge-ei.cz); Eva Guinan, Dana-Farber Cancer Institute and Harvard Medical School (Eva_Guinan@dfci.harvard.edu); Karim R. Lakhani, Harvard Business School & Harvard Institute for Quantitative Social Science (k@hbs.edu). We thank Harvard Catalyst and the Innovation and Implementation (I²) and Advanced Imaging programs for support and cooperation in implementing the experiment, particularly Lee Nadler, William Chin, Laura Weisel, Tom Brady, Tony Hollenberg, and David Frank. Amy Webber and Wei Zhong provided valuable assistance throughout the implementation process. Michael Menietti and Chris Riedl assisted with the execution of the experiment. The Harvard Innovation Laboratory generously provided us with meeting space for the experiment. Support of the Harvard-NASA Tournament Laboratory is acknowledged.

In many fields of science, teamwork has become the dominant mode of knowledge production. Since the 1950s, team size on scientific papers has almost doubled, increasing from 1.9 to 3.5 authors per paper (Wuchty, Jones and Uzzi 2007). In 2009, close to 80% of articles published in the top 4 medical journals were coauthored.¹ Collaboration is driven by the need to combine knowledge, expertise or capabilities embodied in different individuals (Jones 2009) or to access data, equipment and other resources necessary to carry out research. Despite considerable research now being devoted to teams and collaborations formed to perform R&D, relatively little is known about how team members search for and match with one another. These processes of team formation may be particularly important in academic research, where one of the most conspicuous freedoms is the freedom to choose with whom one works.

The process of forming collaborations might be thought of in two stages: individuals first find potential collaborators with complementary knowledge or resources, and then they undertake the decision to enter collaborations. We should observe the formation of a particular collaboration if the expected benefits exceed the costs for each of the parties involved. Thus, information should play an especially important role in the matching process: information is needed both about potential collaborators and about the costs and benefits associated with the collaboration. As in other markets in which agents match, the market for scientific collaborators may also be characterized by search frictions (Mortensen and Pissarides 1999). For example, frictions may arise due to search costs associated with finding potential collaborators, or due to asymmetric information about the ability of potential coauthors or about the quality of the match (Fafchamps, Goyal and van der Leij 2009). We might expect any number of factors to play a role in inducing individuals to choose to collaborate and to successfully coordinate in doing so, and many of these factors are not easily observed to both parties *ex ante*—such as current research interests, personal chemistry and disposition, and timing and scheduling constraints. In fact, many of these factors that may matter in choosing collaborators might only be observed through information gained in close geographic proximity, i.e. through face-to-face interactions (Azoulay, Liu and Stuart 2009).

The broader economic literature has pointed to the importance of geographic and social proximity in explaining information flows between individuals, in firms and in markets (Allen

¹ *New England Journal of Medicine, Lancet, Journal of the American Medical Association (JAMA), and Nature Medicine.*

1970; Cowgill, Wolfers, Zitzewitz, 2008). Consistent with this literature, the existing empirical evidence on scientific collaboration point to the importance of geography and social ties in facilitating collaborations. Several studies have focused on the role of geography, with collaborators tending to be located more geographically proximate (Katz 1994, Mairesse and Turner 2005). Others have shown that scientific teams have become more geographically dispersed in the past 20 years and increasingly span multiple universities (Adams, Clemmons, Black and Stephan 2005, Jones, Wuchty and Uzzi 2008), with evidence suggesting that decreased collaboration costs, particularly the spread of the Internet, has diminished the role of geography (Agrawal and Goldfarb 2008). Research has also pointed to the role of preexisting social ties, such as previous collaborations, in reducing asymmetric information among potential collaborators (Fafchamps, Goyal and Van der Leij 2010) as well as preferences for working with certain kinds of individuals, such as women preferring to work with other women (Boschini and Sjogren 2007).

However, these studies are all based on evidence from collaborations that have already formed, and thus remain mostly silent on the mechanisms by which team members find each other and decide to form collaborations. Consider, for instance, that collaborators tend to be geographically proximate. Is it because geographically close scientists are more likely to run into each other at the proverbial water cooler or elsewhere, exchange ideas and discover grounds for collaboration? Or is it because geographic proximity lowers the costs of collaborating, increasing the returns to the collaboration?

In this paper, we design and execute a field experiment to shed light on the role that colocation and face-to-face interactions can play in reducing frictions in the formation of scientific collaborations. We go further than the existing studies by isolating the role of information gained through colocation from other mechanisms that might play a role in increasing the incidence of collaborations with geographic proximity.² Our experimental setting is a funding opportunity for clinical and imaging researchers at Harvard University and Harvard Medical School's system of hospitals and research centers. Harvard Catalyst, the Harvard Clinical and Translational Center, provides seed funding for clinical and translational research in the form of collaborative pilot grants to faculty and researchers, primarily at Harvard Medical School and its affiliated

² However, given our experimental design, we are not in a position to further disentangle the relative importance of different types of information gained through colocation that can influence the likelihood of collaboration.

institutions. As part of the grant program, applicants were required to participate in an interactive research symposium—a structured face-to-face interaction. The aim of the symposium was to be a forum for investigators to exchange ideas, learn about imaging tools and technologies and meet other researchers. The symposium consisted of a 30-minute general introduction followed by 90 minutes of separate breakout sessions of 20-40 researchers that centered on interactive poster presentations in which each individual researcher effectively “broadcast” to all other participants his or her research project, so as to promote and one-on-one conversations.

It is during this research symposium that we introduce exogenous colocation. Crucial to our research design, the symposium was structured in a way that all participants were exposed to identical “priming” in a 30-minute general session, but then were randomly assigned to the separate and independent 90-minute breakout sessions. Thus, we are able to observe and compare the outcomes of pairs of individuals who participated in the same breakout session in contrast to pairs of individuals who did not - treated and control groups (of pairs) only differed in whether they were colocated during the event in the same breakout room. We then estimate the effect of being colocated in the same breakout room on the likelihood of collaboration, or appearing as co-investigators on a final grant application. We interpret same room location at the event as facilitating face-to-face interactions and thus increasing information flows that may be associated with frictions in the search for collaborators.

We find that being in the same room at the event increases the likelihood of collaboration by about 70% (from 0.16% to 0.28%), although the increase in the number of collaborations is relatively small given that the baseline likelihood of any two researchers collaborating is low. We next investigate for which type of pairs the effect of colocation is greatest. We do so by interacting same room with pair characteristics while controlling for the main effect of being in the same room at the event. We find significant interactions between (1) same room and working in the same clinical area (2) same room and pairs including a woman, and (3) same room and having published together in the past.

The contribution of this paper is twofold. On a substantial level, our results suggest that matching between scientists is subject to considerable frictions and that face-to-face interactions play a central role in the initiation of new collaborations. On a methodological level, we are – to the best of our knowledge - the first to bring field experimental methods to a workplace setting where the participants are engaged in scientific knowledge production. Evidence from

randomized experiments on the scientific community such as ours will be presumably be increasingly valuable to policymakers as they consider reforms to the scientific institutions (Azoulay 2012).

The paper proceeds as follows. We first describe our experimental design in detail, including background about the grant program and details about the symposium in Section I. In Section II we describe the data. The empirical strategy and results follow in Section III and IV respectively. Section V concludes.

I. The Colocation Field Experiment

A. Background: Harvard Medical School and its affiliated hospitals

Our field experiment involved faculty and researchers from Harvard University and its affiliated hospitals and institutions. Harvard Medical School and its 17 affiliated hospitals and research institutes (including Massachusetts General Hospital, Children’s Hospital Boston, Brigham and Women’s Hospital, Beth Israel Deaconess Medical Center and the Dana Farber Cancer Institute) are a major force in biomedical research. Collectively, they employ more than 11,000 faculty and receive in excess of \$1.5 billion in NIH funding per year. Harvard researchers account for around 5% of scientific articles published in the top four medical journals, a larger share than Germany or Canada as a whole.³ Nine Nobel prizes have been awarded for work done at Harvard Medical School.

While the setting of our experiment is based entirely within the Harvard University system, in fact the researchers are working in very distinct institutions. The Harvard affiliated hospitals are separately owned and managed and thus have considerable independence from each other, including separate intellectual property arrangements. They appear as separate entities in hospital rankings and lists of NIH recipients. Four of the five largest hospitals are located on the Longwood Medical Area campus in Boston while Massachusetts General Hospital has its own campus about 3 miles away (see Figure 1 for a map showing the locations of the largest Harvard-affiliated hospitals).

³ Journals included are the *New England Journal of Medicine*, *Journal of the American Medical Association (JAMA)*, *Nature Medicine* and *Lancet*. Authors’ calculations based upon research articles published during the period 2000-2009. Fractional counting was used when coauthors belonged to different institutions.

B. Background: Harvard Catalyst and Advanced Imaging

As part of Harvard's efforts to promote clinical and translational research, Harvard Catalyst, the Harvard Clinical and Translational Center, provides seed funding for clinical and translational research in the form of pilot grants. These pilot grants, in the amount of \$50,000 each, are awarded competitively to faculty within Harvard University. They emphasize early-stage collaborative research with the potential to improve human health.

Our field experiment was layered onto a Harvard Catalyst pilot grant process. This particular process was centered on proposals to move forward the state of knowledge about the applications of advanced imaging technologies to address areas of unmet clinical need. A major challenge in the field of advanced imaging is that progress requires both expertise in the latest imaging tools and technologies and a deep understanding of the health problems to which they could be applied, with these different types of knowledge typically being held by people with different disciplinary backgrounds. Thus, advanced imaging is an archetypical example of a problem often found in modern science: moving the knowledge frontier requires combining knowledge embodied in different individuals.

While the grant process was primarily focused on identifying and funding promising early-stage research in the field of advanced imaging, there was also a perceived need for community building and for familiarizing clinicians with recent developments in advanced imaging. Thus, attendance at an interactive research symposium was included as a required part of the grant process. It was foreseen that these events would be a forum for investigators to exchange ideas, learn about technologies and meet potential collaborators.

In November 2011, all Harvard life science faculty and researchers were invited to participate in a unique funding opportunity centered on advanced imaging (specifically Physiological MR, PET, and Optical Imaging). A total of up to \$800,000 was available to support 15 pilot grants as well as several concept development prizes of \$2,000 each.

In the first stage, investigators who were interested in applying for the grants were asked to submit a Statement of Interest (SOI) in which they briefly described a specific medical problem that advanced imaging techniques could potentially address. Basic biographical information (e.g. degree, institution, department appointment) was also collected at this stage, as well as information about their level of expertise and familiarity with the various health areas and imaging technologies.

The communication about the funding opportunity specified that eligibility to submit a final application was conditional on attending an advanced imaging symposium at pre-announced dates. It also indicated that these events would be studied by Harvard Catalyst to develop better insights about scientific team formation and that data on interaction patterns amongst individuals would be collected.

C. Randomization and the Advanced Imaging Symposium

Four hundred and thirty five applicants were invited to attend the advanced imaging symposium and thus proceed in the grant application process.⁴ Fifty (11.5%) failed to RSVP or otherwise did not show up at the event. Additionally, invitations to attend were extended to several individuals with special expertise in advanced imaging, bringing the total number of participants to 402. The advanced imaging symposium was structured so that participants would come to the event prepared to discuss their idea with other participants in small breakout rooms of 30 to 40 people. The treatment was intended to introduce exogenous colocation to some pairs of participants at the symposium by having them be present in the same breakout rooms at the event. So that a random subset of all possible pairs among all participants would receive the treatment, each participant was randomly allocated to one of the breakout rooms in advance. The event was held over 3 nights with 4 breakout rooms per night.

The events were held January 31, February 1, and February 2, 2012 at the Harvard Innovation Lab located on Harvard's Allston campus. The program began with a short 30-minute address by the program leadership describing the pilot grant opportunity and the agenda for the evening, with a short introduction to imaging tools and technologies. Then, the breakout sessions began in separate rooms on the second floor of the Harvard Innovation Lab. The number of participants in each room varied from 28 to 43.

The breakout room sessions were split into two parts, each 45 minutes long with a 15-minute break in the middle. Standardized posters describing each participant's submitted idea from the SOI (based on information they had provided) were placed in the breakout rooms in advance. Participants were split in two groups: participants from group 1 were asked to stand by their poster during the first session and then during the second session circulate among the other

⁴ Thirty-six statements of interests were outside the parameters of the request for applications in terms of area of inquiry and the submitters were not invited to attend the symposium.

participants' posters, while participants from group 2 did the opposite (i.e. circulated around the room during session 1 and stood near their posters for session 2). Each individual's exact poster placement in the room and their assignment to group 1 or group 2 was randomly determined in advance.

D. Grant Applications

Shortly after the event, symposium participants received via email a full list of the attendees and an invitation to submit applications for the pilot grants or concept awards by the deadline of March 8, 2012. As was typical for previous Harvard Catalyst pilot grant processes, applications had to include a principal investigator and at least one co-investigator. Concept award applications similarly had to include at least 2 applicants. Researchers with faculty appointments could apply as principal investigator on the a pilot grant, but on only one grant application, and could apply as co-investigator on an unlimited number of additional applications. Researchers without a faculty appointment could not be principal investigators on a pilot grant application, but they could be co-investigators on an unlimited number of applications. All attendees were eligible to apply for a concept award grant and could appear on an unlimited number of applications. Finally, at least one co-applicant on any grant application had to have attended the symposium. The grant application did not need to be based upon the initial statement of interest.

Participants were explicitly told before and during the symposium that applications meeting these requirements would be reviewed under a double blind system. They were also told that the composition of the team would not be communicated to reviewers and would not be considered as a criterion for awarding the grant. Thus, participants were not incentivized to collaborate with other symposium participants. In fact, the majority of participants chose not to apply with other symposium participants.⁵

II. Data

A. Sources

We use a variety of data sources to create a dyad-level dataset that will allow us to examine the impact of colocation on collaboration.

⁵ 66% of the applications included only 1 symposium participant as a co-applicant.

Registration data. Faculty and researchers interested in taking part in the funding opportunity were requested to submit a short statement of interest describing in 250 words or less a specific medical problem that advanced imaging techniques could potentially address. At this stage we also collected basic biographical information (rank, education history, hospital affiliation, department), and asked whether they were primarily an imager or a clinician. Clinical area and imaging modality were coded from the SOIs.

Publications. We matched participants to Harvard Catalyst Profiles, a Harvard Medical School database that includes individual publication records for faculty and researchers. From the publication records, we deduced whether scientist pairs in our sample were previous coauthors.

Grant applications. Our main outcome variable comes from the pilot grant and concept award applications. Two hundred and twenty four applications for pilot grants or concept awards were received.⁶ Of those, 148 included one symposium participants in the applicant list, 49 included two symposium participants and 27 included more than two symposium participants. We measure collaboration as any pairs of symposium participants appearing on the same application.

B. Summary Statistics and Randomization Check

Table 1 provides individual-level summary statistics for symposium participants.⁷ Out of 402 attendees, 29% of attendees were females, 42% identified themselves as imagers and 73% of attendees held Harvard faculty appointments (the others being postdoctoral fellows or clinical fellows). Over 80% of attendees came from the four largest Harvard-affiliated hospitals – Massachusetts General Hospital, Brigham and Women’s Hospital, Children’s Hospital Boston and the Beth Israel Deaconess Medical Center. In terms of the clinical areas targeted in the statement of interests, the most popular were neurology (25%), oncology (25%) and neuropsychiatry (10%).

[Insert Table 1 Here]

⁶ 78% of applications were for pilot grants and the rest for concept awards.

⁷ We had 394 individuals attending the events across the 3 nights. However, five individuals with special expertise in advanced imaging attended the event on more than one night, and we count them as different participants on each night, bringing the total number of participants to 402.

Figure 2 shows the distribution of geographic distance between pairs of participants, which we created by geocoding the exact location of their offices and calculating the distance in miles. The distribution clearly has two peaks, with over 30 percent of the pairs located within 0.5 miles of each other and 25 percent of pairs located between 2.5 and 3.0 miles of each other. The first peak corresponds to pairs where both members are based either in the same hospital or in different hospitals from the Longwood Medical Area. The second peak corresponds to pairs where one member is at the Longwood Medical area and the other is at Massachusetts General Hospital.

[Insert Figure 2 Here]

To verify that the randomization generated balance across covariates, we present summary statistics for the pairs in our sample colocated at the event by being assigned to the same breakout room, and those not colocated, in Table 2. As expected given the random assignment, treated pairs and control pairs look very similar.

[Insert Table 2 Here]

In the last row of column 2, we include our outcome variable, collaboration. The incidence of collaboration is noticeably larger in the treated group, which we investigate in a regression framework in the next section. It is notable that the incidence of collaboration is less than 0.2% in our sample. While this may seem low, the likelihood that two HMS faculty members will co-publish in a given year is 0.06% and thus of the same order of magnitude.⁸ Viewed through the lens of all pairwise combinations of scientists who could collaborate, collaboration is indeed a rare event.

[Insert Table 3 Here]

Table 3 shows characteristics of the collaborating dyads. Among attendees who attended on the same night but were not in the same breakout room, there were 33 pairs that co-applied.

⁸ Authors' calculation based upon publication data from Harvard Catalyst Profiles.

Among pairs in the same room at the event, there were 19 pairs that co-applied. It is important to note that some of the within-room collaborations would have occurred in the absence of any treatment effect. Extrapolating the across-room incidence rate (0.16%) to the number of within-room pairs (7,149), we would expect 11 collaborations to have occurred within rooms in the absence of any treatment effect.

III. Empirical strategy

A. Specifications

To estimate the impact of colocation on the likelihood of collaboration between pairs, we run linear regressions with the following specification:

$$Collaboration_{ij} = \alpha + \beta Colocation_{ij} + \delta X_{ij} + \varepsilon_{ij} \quad (1)$$

where the key explanatory variable, $Colocation_{ij}$, is an indicator variable that equals 1 if both researcher i and j were in the same breakout room at the event.⁹ $Collaboration_{ij}$ is an indicator variable for whether i and j appeared on any common pilot grant or concept award applications. X_{ij} is a vector of observable pair-level characteristics that can impact the likelihood of collaboration, including gender composition, differences in rank as well as geographic, scientific and social distance, described below. We estimate equation (1) using a linear probability regression (OLS) with fixed effects for the night of the symposium and heteroskedasticity-robust standard errors.

We might expect the information that colocation provided potential collaborators at the event to be more valuable to some types of pairs than others. For example, it may be the case that individuals from institutions more geographically distant who have fewer other opportunities to meet would benefit more from the lowered search costs and would be more likely to collaborate after meeting at the event. Or it may be that those pairs more geographically proximate in terms of their workplace location will be more likely to collaborate after colocation, since their costs of collaboration are lower and the colocation provides them with additional information about the

⁹ We could also define colocation more broadly as attending the event on the same night. Furthermore, since participants' posters were also randomized within the breakout rooms, colocation could also be defined more narrowly as being neighbors in the breakout room at the event. However, neither of these measures of colocation had a significant impact on our outcome of interest, grant coapplications.

potential collaboration. Similar reasoning follows for other measures of scientific and social distance.

Thus, we next examine whether there was heterogeneity in the effect of being in the same room on collaboration by interacting the colocation indicator with pair-level variables that indicate different measures of social, geographic and scientific distance between pairs with the following regression:

$$Collaboration_{ij} = \alpha + \beta Colocation_{ij} + \theta Colocation_{ij} * Distance_{ij} + \pi Distance_{ij} + \delta X_{ij} + \varepsilon_{ij} \quad (2)$$

where $Distance_{ij}$ is a measure of social, geographic, or scientific distance. The coefficients on the interactions of Colocation with pair characteristics will show how the impact of colocation varies across different measures of distance between the pairs while controlling for the main effect of being colocated.

Investigating the heterogeneity of the effect of being in the same room is associated with two empirical challenges. First, the main effect of colocation on collaboration was driven by a relatively small number of actual within-room collaborations. Moreover, interaction effects between colocation and pair characteristics are driven by even smaller subsets of these few within-room collaborations. Second, pair characteristics may be correlated with one another, which make their interpretation more difficult.

We address with the first issue by relying on conventional significance levels in the pair-level analysis and with the second by introducing the interactions simultaneously rather than considering them separately. However, given the aforementioned empirical challenges, the results from the second specification should be treated with appropriate caution.

B. Covariates

In X_{ij} , we include variables for the gender composition of the pair, including indicators for *Both female*, *One female*, and *Both male*. Gender may matter for collaboration, as evidence shows that women may prefer to work with other women (Boschini and Sjogren 2007). The literature also suggests that female scientists may have more limited academic networks which impacts whom they choose to collaborate with, with the diversity of their networks limited by

family obligations that can prevent attendance at conferences and seminars and due to lower job mobility (see Ding, Levin, Stephan and Winkler, 2010).

For rank, we include indicators for *One postdoc* in the pair and *Both postdocs*. While postdocs were eligible to apply for either the concept or pilot grants, two postdocs could collaborate on a pilot grant application only if a third team member with a faculty appointment assumed the role of principal investigator. We also created indicators for the distance between the ranks of the participants (Postdoc, Instructor, Assistant Professor, Associate Professor, and Professor).

We draw upon various measures to capture the geographic distance between pairs. As noted earlier, the empirical evidence on collaboration shows that collaborators tend to be located more geographically proximate (Katz 1994, Mairesse and Turner 2005). We create an indicator for *Same hospital*, which indicates whether their primary appointment is in the same Harvard-affiliated hospital or institutes. The Harvard Medical School hospitals and institutes in our sample are located throughout the Boston area, however the largest concentration of researchers are located in hospitals and institutes either on the Longwood Medical Area (LMA) campus or at the Massachusetts General Hospital (MGH) campus. The campuses are located approximately 3 miles apart (taking about 20 minutes of travel during normal traffic flows). To capture this feature of the institutional environment, we also create an indicator for *Both Longwood Campus*, indicating that both members of the pair were on same campus.¹⁰ We also create a direct measure of geographic distance between the pairs by geocoding the exact location of their offices and calculating the distance in miles.

For scientific distance between the pair members, we first create variables to capture the complementarity of knowledge and skills that may exist for clinicians working with imagers. Thus, we create indicators for *Both imagers*, *One imager-One clinician*, and *Both clinicians*. We used information that attendees provided during the initial stage of the application process to construct this variable, in which they self-identified as being primarily an imager or a clinician. Next, we created indicators for the *Same clinical area* and *Same imaging modality* (Physiological MR, PET, or Optical Imaging), which were coded from the SOIs submitted in the first stage of the application process. We also create measures of scientific distance using overlap in the

¹⁰ The LMA includes 8 hospitals/institutes in our sample and the MGH campus includes 2 hospitals/institutes. The other hospitals/institutes in the sample are considered to be individual campuses.

MeSH terms from each individual’s publications, and overlaps in the keywords of each individual’s SOI.

Finally, we incorporate information on whether the pair had previously collaborated to measure social distance beyond scientific distance. We create an indicator for *Previous coauthors*, and we further capture the intensity of the collaboration by creating an indicator for collaborations who only had *1 Copublication* in the past and those who had *More than 1 copublication*.

IV. Results

A. Does colocation increase the propensity to collaborate?

Our results on the main effect of colocation at the event on collaboration are presented in Table 4. Column 1 shows the basic result of regressing whether the pair was in the same breakout room at the event on whether they submitted a joint grant or concept award application. Colocation increases the likelihood of collaborating on an application by approximately 70% (increasing the likelihood of a pair collaborating from 0.16% to 0.28%). The estimate is significant at the 10% confidence level.

[Insert Table 4 Here]

The advanced imaging symposia were held on three different nights. We thus include fixed effects for the night of the event (January 31, February 1, or February 2) in Column 2 to account for any differences across nights. The night fixed effects are not significant and their inclusion has very little impact on the same room coefficient (or its standard error).

The random assignment ensures that being in the same room is orthogonal asymptotically to any observable or unobservable pair characteristic.¹¹ However, the direction and magnitudes of these variables allow us to confirm whether the impact of various factors on the likelihood of

¹¹ In finite samples, colocation could be correlated with pair characteristics by chance. While this is much less of a concern than in observational data (Leamer 2009), it is nonetheless useful to control for relevant observable pair characteristics to address the possibility that the effect of colocation is affected by differences in observable pair characteristics. Introducing controls has the added benefit of improving the precision of the colocation estimate by reducing the unexplained variance.

collaboration in our context are consistent with the existing literature on collaboration, and allow us to contrast the magnitude of the effect of our same room treatment with these other factors.

In Columns 3 and 4 we thus introduce pair-level variables to account for gender composition, differences in rank as well as geographic, scientific and social distance. The coefficients on the additional pair-level variables show that working in the same clinical area, being affiliated with the same hospital and being a coauthor in the past are positively and significantly correlated with collaboration. Consistent with the related literature, these results suggest that geographic, scientific and social proximity are all positively related to collaboration. Given the likely complementarities of skills and knowledge between imagers and clinicians, our prior was that collaborations would be more likely to form when one pair member was a clinician and the other was an imager. However, the results show that collaborations were more likely to form when both members of the pair were imagers. Collaboration was significantly less likely to occur between pairs consisting of two postdocs, which is possibly explained by the fact that two postdocs could collaborate on a pilot grant application only if a third team member with a faculty appointment assumed the role of principal investigator.

In the specifications controlling for these other factors that impact collaboration, the point estimate for the effect of being in the same room increases slightly from 0.0012 to 0.0014 and is significant at the 5% confidence level in Column 4, which accounts for the intensity of previous collaborations. Relative to the effects of geographic, scientific and social distance on collaboration, our estimated effect of colocation on collaboration is over 30% of the effect of being from the same hospital (0.0044) and of researching the same clinical area (0.0040), but only about 1% of the effect of having coauthored in the past (0.1126).

These results suggest that frictions do exist in the process by which collaborations form. If no frictions existed, then being colocated at the event should have no impact, and the coefficient on same room would not have been significant. Thus, on average, being colocated at the advanced imaging symposium, and the face-to-face interactions that this colocation enabled, reduced search costs or asymmetric information that would have otherwise prevented some collaborations from being realized. Relative to the other factors that impact scientific collaboration in our sample, we find that the effect colocation at the event was relatively large compared to the effect of geographic and scientific proximity, but still small compared to social proximity.

B. For which pairs does colocation have a larger effect?

We investigate in Table 5 whether colocation has a differential effect for different types of pairs. We introduce the interactions simultaneously as covariate characteristics are often correlated with each other. In the first specification in the first column, we see that if the pair is working on a topic in the same clinical area, the effect of colocation is positive, increasing the likelihood of a pair collaborating from 0.35% (the sample average incidence of collaboration for pairs working in the same clinical area but not colocated) to 0.94%. This result shows that researchers more proximate in scientific space were more likely to convert their interactions at the event into a collaboration. There are several possible explanations for the effect, but it suggests that researchers had limited information about these potential collaborations – either about who else was working on applying advanced imaging to the same clinical area, or about the potential benefits of collaborations with these individuals. If they did, the information that we provided through colocation should not have any independent effect for these pairs. It may also be the case that face-to-face discussions were more beneficial given the common ground they shared, which allowed them to convert the discussions to collaborations. Another possible explanation is that it is costly to switch clinical areas, therefore, even if researchers talked to people with interesting ideas in other areas at the event, the benefits to collaboration were highest for those in the same clinical area.

[Insert Table 5 Here]

Next, we find a large and positive effect of being in the same room and having coauthored a paper in the past, which we consider as a proxy for social proximity. In the second column, we decompose previous co-authorship into pairs that had coauthored exactly one paper and pairs that had coauthored more than one paper. The results show that being a more frequent coauthor (having published more than one publication together) increases the likelihood of collaborating after being colocated at the event quite dramatically. These results suggest that interacting at the event reconnected such existing but perhaps “dormant” ties. The social network literature suggests that reconnecting dormant ties can be a useful source of knowledge and social capital (Levin, Walter and Murnighan 2011), which is consistent with our findings. In addition, given that people trust their coauthors, the event might have provided an opportunity to discuss new

ideas with previous coauthors in the room that they were not as likely to discuss with others. Similar to the reasoning with scientific proximity, these face-to-face discussions may have then been more beneficial, which allowed them to convert the discussions to collaborations.

Finally, we also see that gender played a role in the pairs that collaborated after being colocated after the event. If there was one woman in the pair, being in the same room increases the likelihood of collaboration by 161% over the average collaboration rate in the sample. Being in the same room also seems to have an impact of pairs consisting of two women, although the effect is not significant given that there was a small number of such pairs. These results indicate that there tended to be a greater return to the information about potential collaborators gained at the event when the pair included one or two women. These results are consistent with Ding et al.'s (2010) findings that IT benefited collaborations more for female scientists than for male scientists, given that women have less diverse networks, have lower job mobility, and more constraints to attending conferences and seminars. These factors would lead women to similarly benefit more from colocation with other researchers at the event in terms of finding coauthors.

Our results on the interaction between colocation at the event and geographic distance are not conclusive. The point estimates for the interaction of same room and same hospital is positive but not significant and the same holds for the interaction of same room and both on the Longwood campus.

We experimented with various alternative specifications such as including more fine-grained measure of geographic distance, scientific distance or social distance¹² as well as controlling more flexibly for ranks and rank differences between pair members (for alternative treatment of geographic distance see Appendix Tables 1 and 2; other alternative specifications are not reported). We have relatively few pairs that are close in social space and this causes the interaction of social proximity and same room to lose significance in some specifications with multiple measures of social proximity. Our other interaction results, however, are consistent and stable across specifications.

¹² We considered for instance whether pair members investigated the same imaging modality, the extent of the overlap of scientific keywords in previous publications and whether pair members shared a common coauthor.

V. Conclusions

In this paper, we showed that temporary colocation increases the likelihood of collaboration between scientists and that pairs close in scientific and social place or including women were more strongly impacted. We interpret these results as showing that face-to-face interactions can increase information flows that may be associated with frictions in the search for collaborators. Overall, our findings suggest that search costs and instances of asymmetric information powerfully shape the type of collaborations that occur between scientists and can lead to frictions in the formation of collaborations.

Modern biomedical research is characterized by heavy concentrations of researchers in very specific areas and by a multiplicity of conferences and symposia that bring together geographically distant scientists. Both phenomena could, at least in part, be explained by the mechanisms investigated in this paper. We find that face-to-face interactions play an important role in the initiation of collaborations and such interactions require either temporary or permanent colocation.

Field experiments often raise questions of external validity and ours is no exception. All scientists pairs involved in our experiment were from a single academic system – Harvard University and its affiliated hospitals – which in addition is quite different from the average biomedical research center in several important dimensions including size. Our field experiment did not involve pairs of scientists whose offices or labs are more than a cab ride away. Thus, our results may be less informative regarding matching between scientists that occurs at conferences. Another limitation is that participants of the symposium were all focused on a particular form of scientific inquiry - the application of advanced imaging technologies and tools to improve clinical outcomes. It could be that this specific field is different from others in ways that could influence the results.

Given the small sample size of collaborations and as it is too early to observe longer-run outcomes of scientific productivity, such as subsequent publications, we are not in a position yet to assess whether the collaborations within the same rooms (and thus treated by random colocation) were better, worse, or different than collaborations across rooms. For example, in contemporary work drawing upon a natural experiment, Catalini (2012) finds that random versus chosen co-location leads to higher levels of experimentation (variance of citations to papers).

Despite these limitations, we see the present study as a step in opening the black box of how scientific collaborations form. Our study is the first to show that matching between scientists is subject to considerable frictions and that face-to-face interactions play a central role in the initiation of new collaborations. On a methodological level, we are – to the best of our knowledge - the first to bring field experimental methods to a workplace setting in the scientific community.

In recent years there has been considerable interest in the policy arena in fostering collaborative and especially interdisciplinary collaborations. Yet there is scant evidence on how to do this in practice. Here we show that creating settings where scientists meet face-to-face and specifically discuss early-stage research ideas can be useful for fostering collaboration. However, time spent in such “mixer” events has opportunity costs and we thus remain agnostic on the effect of such activities on scientific productivity and welfare more generally.

REFERENCES

- Adams J, Clemmons R, Black C and Stephan P (2005). "Scientific Teams and Institutional Collaborations: Evidence from U.S. Universities," *Research Policy* 34:259-285.
- Agrawal A, Goldfarb A (2008). "Restructuring Research: Communication Costs and the Democratization of University Innovation," *American Economic Review*, 98(4):1578-1590.
- Allen T (1970). "Communication Networks in R&D Laboratories," *R&D Management* 1(1):14-21.
- Azoulay P (2012). "Research efficiency: Turn the scientific method on ourselves," *Nature*, 484, 31–32 (05 April 2012).
- Azoulay P, Liu C, Stuart T (2009). "Social Influence Given (Partially) Deliberate Matching: Career Imprints in the Creation of Academic Entrepreneurs" HBS Working Paper 9-136.
- Boschini A, Sjogren A (2007). "Is Team Formation Gender Neutral? Evidence from Coauthorship Patterns," *Journal of Labor Economics*, 25(2):325-365.
- Catalini C (2012). "Microgeography and the Direction of Inventive Activity" mimeo. Rotman School of Management, University of Toronto.
- Cowgill B, Wolfers J and Zitzewitz E (2008). "Using Prediction Markets to Track Information Flows: Evidence from Google" mimeo, Dartmouth University.
- Ding W., Levin S., Stephan P., and Winkler A. (2010). The Impact of Information Technology on Academic Scientists' Productivity and Collaboration Patterns," *Management Science*, Vol. 56, No. 9, September 2010, pp. 1439–1461.
- Fafchamps M, Goyal S and Van der Leij M (2010). "Matching and Network Effects," *Journal of the European Economic Association*, 8(1):203-231.
- Katz J (1994). "Geographical proximity and scientific collaboration," *Scientometrics* 31(1):31-43.
- Leamer E (2010). "Tantalus on the road to Asymptotia," *Journal of Economic Perspectives* 24(2):31–46.
- Levin D, Walter J and Murnighan, J. (2011). "Dormant Ties: The Value of Reconnecting." *Organization Science*, Vol. 22, No. 4, pp. 923-939, 2011.
- Jones B (2009). "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?" *Review of Economic Studies* 76:283-317.

Jones B, Wuchty S, and Uzzi B (2008) "Multi-university Research Teams: Shifting Impact, Geography, and Stratification in Science," *Science*, 322: 1259.

Mairesse J, Turner L (2005). "Measurement and explanation of the intensity of co-publication in science: an analysis at the laboratory level" NBER Working Paper 11172.

Mortensen, D, Pissarides, C (1999). "New developments in models of search in the labor market", In: Orley C. Ashenfelter and David Card, Editor(s), *Handbook of Labor Economics*, Elsevier, 1999, Volume 3, Part B, Pages 2567-2627.

Wuchty S, Jones B, and Uzzi B (2007). "The Increasing Dominance of Teams in Production of Knowledge," *Science* 316, 1036.

Table 1. Summary Statistics, Attendees

	Sample Mean
Female	0.29
Faculty Member	0.73
Imager	0.42
Longwood Campus	0.51
Hospital	
Massachusetts General Hospital	0.37
Brigham and Women's Hospital	0.19
Beth Israel Deaconess Medical Center	0.14
Children's Hospital Boston	0.13
Other	0.17
Clinical Area (SOI)	
Neurology	0.25
Oncology	0.25
Neuropsychiatric	0.10
Cardiovascular	0.06
Gastroenterology	0.04
Transplantation	0.04
Ophthalmology	0.03
Other	0.23
Attended on Jan. 31	0.35
Attended on Feb. 1	0.32
Attended on Feb. 2	0.33
Observations	402

Table 2. Dyads by Colocation Treatment

Sample Means	<i>Same Room Treatment: Colocated</i>	<i>Control: Not Colocated</i>
One postdoc	0.403	0.396
Both postdocs	0.071	0.074
One female	0.402	0.418
Both female	0.085	0.081
Same Hospital	0.199	0.209
Both Longwood Campus	0.265	0.257
One Imager-One Clinician	0.492	0.490
Both imagers	0.176	0.177
Same Clinical Area (SOI)	0.124	0.119
Previous Coauthor	0.001	0.002
Previous Coauthor (1 Copub)	0.001	0.001
Previous Coauthor (>1 Copubs)	0.000	0.001
Collaboration (Outcome variable)	0.0028	0.0016
Observations	6,730	20,059

Notes: The *Same Room Treatment* was being colocated in the same room at the event and was randomized across pairs of participants attending on the same night. *Collaboration* indicates whether the pair appeared on any common pilot grant or concept award applications. See Section III in the text for a detailed description of the variables.

Table 3. Collaborating Dyads by Colocation Treatment

Sample Means	<i>Collaborations within the same room</i>	<i>Collaborations across rooms</i>
One postdoc	0.421	0.182
Both postdocs	0.000	0.030
One female	0.474	0.303
Both female	0.158	0.061
Same Hospital	0.579	0.667
Both Longwood Campus	0.158	0.303
One Imager-One Clinician	0.474	0.455
Both imagers	0.316	0.424
Same Clinical Area (SOI)	0.579	0.273
Previous Coauthor	0.105	0.121
Previous Coauthor (1 Copub)	0.053	0.000
Previous Coauthor (>1 Copubs)	0.053	0.121
Observations	19	33

Notes: *Collaboration* indicates the pair appeared on a common pilot grant or concept award application. See Section III in the text for a detailed description of the variables.

Table 4. Main effect of Colocation on Collaboration

DV = Collaboration	(1)	(2)	(3)	(4)
Same Room Treatment	0.0012 ⁺ (0.0007)	0.0012 ⁺ (0.0007)	0.0014 ⁺ (0.0007)	0.0014 [*] (0.0007)
One postdoc			-0.0008 (0.0005)	-0.0008 (0.0005)
Both postdocs			-0.0015 [*] (0.0007)	-0.0015 [*] (0.0007)
One female			0.0001 (0.0006)	0.0001 (0.0006)
Both female			0.0010 (0.0011)	0.0009 (0.0011)
Same Hospital			0.0044 ^{**} (0.0010)	0.0044 ^{**} (0.0010)
Both Longwood Campus			-0.0002 (0.0006)	-0.0002 (0.0006)
One Imager-One Clinician			0.0008 ⁺ (0.0005)	0.0009 ⁺ (0.0005)
Both imagers			0.0026 ^{**} (0.0010)	0.0026 ^{**} (0.0010)
Same Clinical Area (SOI)			0.0040 ^{**} (0.0014)	0.0040 ^{**} (0.0014)
Previous Coauthor			0.1126 [*] (0.0451)	
Prev. Coauthor (1 Copub)				0.0333 (0.0377)
Prev. Coauthor (>1 Copubs)				0.1951 [*] (0.0799)
Constant	0.0016 ^{**} (0.0003)	0.0012 ^{**} (0.0004)	-0.0010 (0.0007)	-0.0010 (0.0007)
Night fixed effects	No	Yes	Yes	Yes
R2	0.000	0.000	0.017	0.023
Nb. of Obs.	26,789	26,789	26,789	26,789

Notes: Dependent variable is *Collaboration*, an indicator variable for whether the pair appeared on any common pilot grant or concept award applications. The main variable of interest is *Same Room Treatment*, which was randomized across pairs attending on the same night. All estimation is by OLS. Robust standard errors in parentheses. ⁺ $p < 0.10$, ^{*} $p < 0.05$, ^{**} $p < 0.01$

Table 5. Colocation and Interactions with Measures of Distance

DV = Collaboration	(1)	(2)
Same Room Treatment	-0.0019 (0.0015)	-0.0017 (0.0015)
One postdoc	-0.0013* (0.0005)	-0.0013* (0.0005)
Same Rm X 1 postdoc	0.0019 (0.0015)	0.0019 (0.0015)
Both postdocs	-0.0013 (0.0008)	-0.0012 (0.0008)
Same Rm X Both postdocs	-0.0010 (0.0012)	-0.0010 (0.0012)
One female	-0.0005 (0.0006)	-0.0005 (0.0006)
Same Rm X One female	0.0025 (0.0015)	0.0023 (0.0015)
Both female	-0.0001 (0.0010)	-0.0001 (0.0010)
Same Rm X Both female	0.0036 (0.0030)	0.0022 (0.0027)
Same Hospital	0.0039** (0.0011)	0.0038** (0.0011)
Same Rm X Same Hospital	0.0020 (0.0027)	0.0017 (0.0027)
Both Longwood Campus	0.0003 (0.0007)	0.0002 (0.0007)
Same Rm X Both Longwood	-0.0017 (0.0015)	-0.0015 (0.0015)
One Imager-One Clinician	0.0008 (0.0005)	0.0009+ (0.0005)
Same Rm X One imager	-0.0000 (0.0013)	-0.0001 (0.0013)
Both imagers	0.0027* (0.0011)	0.0028* (0.0011)
Same Rm X Both imager	-0.0004 (0.0024)	-0.0011 (0.0023)
Same Clinical Area (SOI)	0.0016 (0.0013)	0.0016 (0.0013)
Same Rm X Same Clin Area	0.0094* (0.0040)	0.0098* (0.0040)
Previous Coauthor	0.0845* (0.0424)	
Same Rm X Prev Coau	0.2408 (0.1963)	
Previous Coauthor (1 Copub)		-0.0044** (0.0011)
Same Rm X 1 Copub		0.1968 (0.1781)
Previous Coauthor (>1 Copubs)		0.1623* (0.0760)
Same Rm X >1 Copubs		0.8291** (0.0762)
Constant	-0.0001 (0.0006)	-0.0001 (0.0006)
Night fixed effects	Yes	Yes
R2	0.024	0.040
Nb. of Obs.	26,789	26,789

Notes: Dependent variable is *Collaboration*, an indicator variable for whether the pair appeared on any common pilot grant or concept award applications. The main variable of interest is *Same Room Treatment*, which was randomized across pairs attending on the same night. All estimation is by OLS. Robust standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Figure 1. Map of the four largest Harvard Medical School affiliates

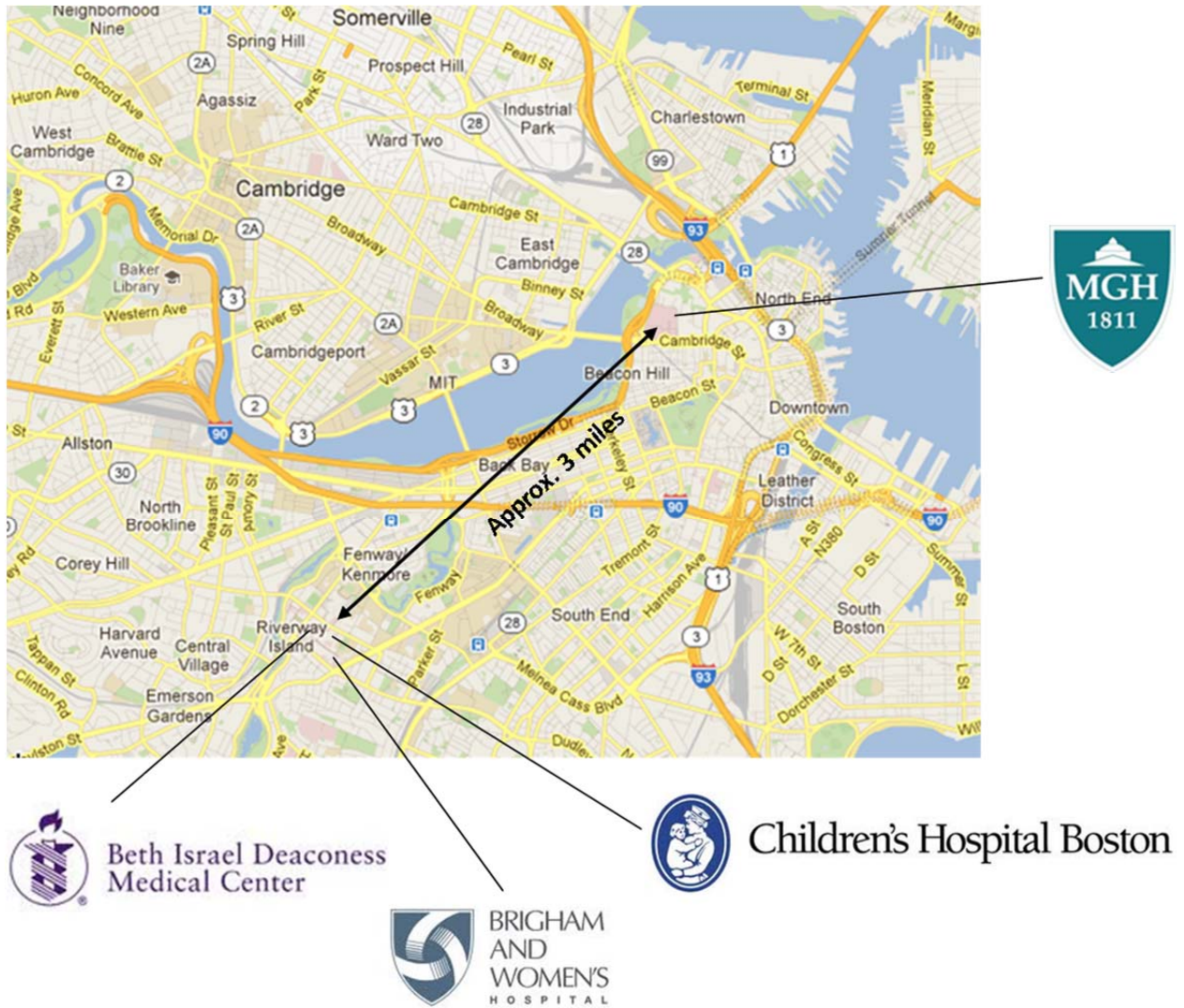
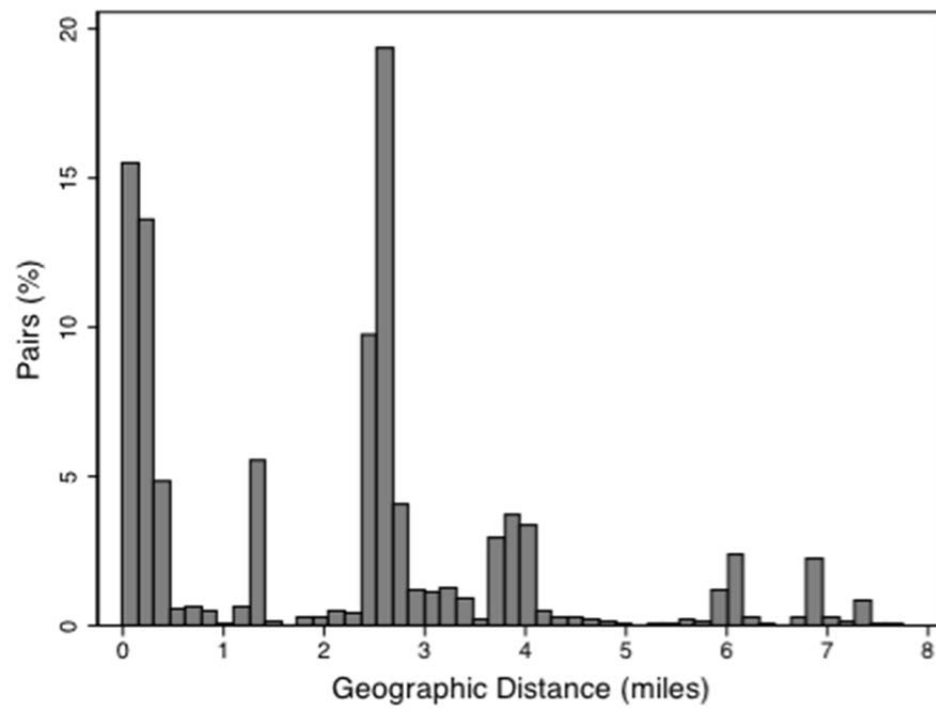


Figure 2. Geographic Distance Between Pairs



Notes: Distance between pairs of researchers in the sample was calculated by geocoding the exact location of their offices and calculating the distance in miles.

Appendix Table 1. Main effect of Colocation on Collaboration, Including Geographic Distance

DV = Collaboration	(1)	(2)
Same Room Treatment	0.0014*	0.0014*
	(0.0007)	(0.0007)
One postdoc	-0.0008	-0.0010 ⁺
	(0.0005)	(0.0005)
Both postdocs	-0.0015*	-0.0018*
	(0.0007)	(0.0007)
One female	0.0001	0.0001
	(0.0006)	(0.0006)
Both female	0.0009	0.0009
	(0.0011)	(0.0011)
Same Hospital	0.0044**	
	(0.0010)	
Both Longwood Campus	-0.0002	
	(0.0006)	
One Imager-One Clinician	0.0009 ⁺	0.0010*
	(0.0005)	(0.0005)
Both imagers	0.0026**	0.0029**
	(0.0010)	(0.0010)
Same Clinical Area (SOI)	0.0040**	0.0041**
	(0.0014)	(0.0014)
Previous Coauthor (1 Copub)	0.0333	0.0337
	(0.0377)	(0.0377)
Previous Coauthor (>1 Copubs)	0.1951*	0.1957*
	(0.0799)	(0.0800)
<i>Geographic Distance</i>		
0.2 to <1 mi (0 to <0.2 mi omitted)		-0.0007
		(0.0012)
1 to <2 mi		-0.0004
		(0.0017)
2+ mi		-0.0028**
		(0.0008)
Constant	-0.0010	0.0016 ⁺
	(0.0007)	(0.0009)
Night fixed effects	Yes	Yes
R2	0.023	0.022
Nb. of Obs.	26,789	26,789

Notes: Dependent variable is Collaboration, an indicator variable for whether the pair appeared on any common pilot grant or concept award applications. The main variable of interest is Same Room Treatment, which was randomized across pairs attending on the same night. All estimation is by OLS. Robust standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Appendix Table 2. Colocation and Interactions, Including Geographic Distance

DV = Collaboration	(1)
Same Room Treatment	-0.0026 (0.0016)
One postdoc	-0.0014* (0.0006)
Same Rm X 1 postdoc	0.0018 (0.0014)
Both postdocs	-0.0016+ (0.0009)
Same Rm X Both postdocs	-0.0009 (0.0012)
One female	-0.0005 (0.0006)
Same Rm X One female	0.0024 (0.0015)
Both female	-0.0000 (0.0010)
Same Rm X Both female	0.0022 (0.0028)
One Imager-One Clinician	0.0011* (0.0005)
Same Rm X One imager	-0.0002 (0.0013)
Both imagers	0.0032** (0.0011)
Same Rm X Both imager	-0.0012 (0.0024)
Same Clinical Area (SOI)	0.0017 (0.0013)
Same Rm X Same Clin Area	0.0099* (0.0041)
Previous Coauthor (1 Copub)	-0.0046** (0.0011)
Same Rm X 1 Copub	0.1989 (0.1777)
Previous Coauthor (>1 Copubs)	0.1625* (0.0761)
Same Rm X >1 Copubs	0.8302** (0.0762)
<i>Geographic Distance</i>	
0.2 to <1 mi (0 to <0.2 mi omitted)	-0.0013 (0.0014)
1 to <2 mi	-0.0019 (0.0016)
2+ mi	-0.0031** (0.0009)
Same Rm X 0.2 to <1 mi	0.0018 (0.0031)
Same Rm X 1 to <2 mi	0.0040 (0.0046)
Same Rm X 2+ mi	0.0007 (0.0020)
Constant	0.0029** (0.0010)
Night fixed effects	Yes
R2	0.040
Nb. of Obs.	26,789

Notes: Dependent variable is Collaboration, an indicator variable for whether the pair appeared on any common pilot grant or concept award applications. The main variable of interest is Same Room Treatment, which was randomized across pairs attending on the same night. All estimation is by OLS. Robust standard errors in parentheses. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$